The prognostic roles of mRNAs of the exosomes derived from bone marrow stromal cells in common malignancies: a bioinformatic study

Fei Zhao
Yan-Qiu Yu
Department of Pathophysiology, College of Basic Medicine Science, China Medical University, Shenyang, Liaoning, China

Background: The inclusion of exosomes enters the recipient cells by means of endocytosis or direct fusion for information exchange between cells and cells. The inclusion of BMSCs-exo helps to guide the diagnosis and prognosis of cancer, especially cancer.

Purpose: This research was to systematically elucidate the prognostic value of mRNAs of the exosomes derived from bone marrow stromal cells (BMSCs) in common malignant neoplasms, such as breast cancer, ovarian cancer, lung cancer, and gastric cancer.

Methods: Gene expression data (GSE78235) for the exosomes derived from BMSCs were extracted from the Gene Expression Omnibus database. Firstly, the differentially expressed genes were detected by comparing the RNA expression from exosomes derived from BMSCs between four tumor patients and two healthy controls using the limma package. Subsequently, functional enrichment analysis, including Gene Ontology and Kyoto Encyclopedia of Genes and Genomes of differentially expressed genes, was performed using the Functional Enrichment analysis tool (FunRich v3.1.3) software followed by the construction of a protein–protein interaction (PPI) network via STRING v3.6.0. Molecular Complex Detection was used to screen the hub proteins by setting up the following threshold score $\geq 4$ and nodes $\geq 10$. Cytoscape v3.6.1 was used to for visualizing PPI network. Finally, Kaplan–Meier analysis for hub proteins was performed by Kaplan–Meier plotter online platform.

Results: A total of 386 genes originating from the exosomes derived from BMSCs were identified as statistically significant ($P < 0.05$, FDR $< 0.05$), which consisted of 150 upregulated genes and 236 downregulated genes. Also, 32 pathways were identified as significant ($P < 0.05$, FDR $< 0.05$). The PPI network of exosomes derived from BMSC proteins included 100 protein nodes with 579 interaction edges. The hub proteins, including PODN, ZNF521, and CFI, which interacted with ten or more other proteins, were indicated as the hub proteins of PPN of exosomes derived from BMSCs.

Conclusion: Taken together, our findings revealed the prognostic roles of mRNAs of exosomes derived from BMSCs and provided implications for targeted therapy for common malignant neoplasms. However, further studies require large samples and experimental verification.

Keywords: bioinformatics, BMSCs, clinical outcome, exosome, malignant neoplasms

Introduction

Mesenchymal stromal cells (MSCs), also known as mesenchymal stem cells, have the potential for self-renewal and multidirectional differentiation.\(^1\) Considering the source, MSCs can be isolated from multiple tissues, such as bone marrow, umbilical cord, blood, placenta, and adipose tissue.\(^2,3\) It was reported that bone marrow mesenchymal cells (BMSCs) that originated from the bone marrow can secrete a diverse variety of bioactive factors in an autocrine and/or paracrine way, which was expected to become
a non-cellular therapy for the treatment of various diseases, especially tumors. Of them, an important paracrine factor of BMSCs is the exosome, which is a lipid membrane wrapped vesicle with a diameter of 40–200 nm. Exosome is abundant in miRNA, mRNA, lipids, and proteins. It has been reported that exosomes enter the recipient cells by endocytosis, or directly fuse with the recipient cells to exchange information between cells. Exosomes derived from BMSCs have been able to treat proangiogenic, cardiovascular disease, liver disease, lung injury, and kidney injury. Indeed, there are two opposite viewpoints on the role of exosome in the treatment of tumors. Some studies have suggested that exosomes have a strong antitumor effect, while other studies have shown that exosomes derived from MSCs can promote the growth and metastasis of cancer cells. Disappointingly, the mechanism of antitumor effect of MSC-derived exosomes remains unclear.

Malignant tumors cause an extremely large burden around the world, second only to cardiovascular disease. Despite using combinations of various strategies of diagnosis and treatment, the prognosis of cancer patients remains poor. Therefore, it is necessary to explore multiple effective early diagnostic and prognostic markers. Recently, it has been reported that exosomes are involved in cancer development, progression, and resistance; therefore, they may be used as multiple targets for the development of a diagnostic tool for early detection of cancer. Along with the advances in microarray and life molecules sequencing technology, high-throughput research methods are widely used in biomedicine. High-throughput and corresponding bioinformatics analysis of exosomes is helpful to guide the diagnosis and prognosis of diseases, especially cancer. In this study, our goal was to identify the prognostic roles of differentially expressed genes originating from exosomes derived from BMSCs in common cancers, which provide new diagnostic and therapeutic targets for cancers.

**Materials and methods**

**Microarray data extraction and identifying differentially expressed genes**

We downloaded the gene expression profile (GSE78235) from the public availability repository Gene Expression Omnibus (GEO; database. GEO2R online tool allows researchers to carry out R-based analysis of GEO data to identify differentially expressed genes (DEGs). Subsequently, the Database for Annotation, Visualization, and Integrated Discovery (https://david.ncifcrf.gov/) v6.8 was used to perform re-annotation for DEGs. A gene is defined as a DEG between the tumor sample and the normal control sample when the false discovery rate (FDR) is <0.05 (FDR ≤ 0.05) and the fold change (FC) is at least two times higher or lower (|log2FC|≥1).

**Functional enrichment analysis for DEGs**

Functional enrichment analysis was performed for DEGs to give a functional overview of the DEGs by calculating the whole significance of the gene expression. Gene ontology (GO) comprises molecular function, cellular component, and biological process. Kyoto Encyclopedia of Genes and Genomes (KEGG) was used to clarify how DEGs perform function through a certain path. We selected Functional Enrichment analysis tool (FunRich v3.1.3), which is designed to handle a variety of gene/protein datasets irrespective of the organism, to perform pathway enrichment analysis. FDR <0.05 was set as the cut-off criterion.

**Construction of protein–protein interaction network for DEGs and identification of hub proteins**

Protein–protein interaction (PPI) was used to analyze the interrelationship between differentially expressed proteins, and further elucidate the mechanisms of genes playing important roles in physiological and pathological conditions. The STRING database (http://string-db.org) helps to provide a critical assessment and integration of protein–protein interactions, including direct (physical) and indirect (functional) associations. In this study, we used STRING database v10.5 to construct PPI network for DEGs. The parameters were set as follows: meaning of network edges: confidence; minimum required interaction score: medium confidence (0.400). Then, Cytoscape software v3.6.0 was used to visualize the PPI network. At last, Cytoscape software plug-in Molecular Complex Detection was used to screen hub proteins within the PPI network.

**Kaplan–Meier (KM) plot analysis for hub proteins**

We ran a KM analysis for three hub genes to characterize the association of gene expression and corresponding clinical outcome in common malignant tumor. KM analysis is available on the website http://www.kmplot.com. The KM plotter is capable to assess the effect of any gene or gene combination on survival in breast, ovarian, lung, gastric, liver cancer patients using over 30,000 samples measured using gene chips or RNA-seq.

**Results**

**Identification of DEGs from BMSCs-derived exosomes**

Based on GEO2R online platform, using FDR<0.05 and (|log2FC|≥1) as the cut-offs, 386 genes were identified to
be differentially expressed in the tumor compared with the normal control samples, which included 150 upregulated genes and 236 downregulated genes.

**GO enrichment analysis**

The eight important terms are obtained by GO enrichment analysis in the software FunRich as shown in Figures 1A–C. The genes were enriched in biological processes of response to signal transduction, cell communication, cell growth and maintenance, and energy pathways metabolism. As for the molecular function, these genes showed enrichment in GTPase activity, chaperone activity, catalytic activity, calcium ion binding, and receptor signaling complex scaffold activity. Besides, the cell component indicated enrichment predominantly at exosomes, cytoplasm, lysosome, nucleus, and plasma membrane.

**KEGG pathway analysis**

KEGG pathway enrichment analysis suggested that the genes were significantly enriched in pathways including vascular endothelial growth factor (VEGF) and VEGF receptor (VEGFR) signaling network, thrombin/protease-activated receptor pathway, PAR1-mediated thrombin signaling events, signaling events mediated by VEGFR1 and VEGFR2, integrin family cell surface interactions, IL5-mediated...
signaling events, beta1 integrin cell surface interactions, syndecan-1-mediated signaling events, plasma membrane estrogen receptor signaling, and proteoglycan-mediated signaling events (Figure 1D).

PPI network construction and hub module analysis

We constructed a network for DEGs from the exosomes derived from BMSCs that included 579 interaction edges and 150 proteins nodes (Figure 2A). Next, we obtained four key modules by means of Cytoscape plug-in Molecular Complex Detection. At last, the three proteins that interacted with at least ten other proteins ($P<0.05$, FDR $<0.05$), which included PODN, ZNF521, and CFI, were identified as the hub proteins from four key modules (Figure 2B–E).

Survival analysis of hub proteins

We ran a KM plot analysis for three genes to visualize the association of hub genes expression and match the clinical outcome in breast cancer, ovarian cancer, lung cancer, and gastric cancer. The upregulation of PODN, ZNF521, and CFI indicated good survival in breast cancer (Figure 3A–C). Downregulation of PODN and ZNF521 indicated a good outcome in ovarian cancer (Figure 3D and E), while the upregulation of CFI indicated a good survival in ovarian cancer (Figure 3L). The downregulation of PODN, ZNF521, and CFI indicated good survival in lung cancer Figure 3G–I), while the upregulation of CFI indicated a good survival in gastric cancer (Figure 3J and K), while the upregulation of CFI indicated a good survival in gastric cancer (Figure 3L).

Discussion

Malignant tumor is a major public health problem that threatens human health and is second only to cardiovascular disease. Unfortunately, tumors lack early diagnostic and prognostic markers with a high sensitivity and specificity. Our study focuses on the critical role of mRNA of exosome derived from BMSCs. In our study, by extracting the gene expression information from GEO database, we obtained 386 DEGs, including 150 upregulated genes and 236 downregulated genes from the exosome derived from BMSCs. Next, we uncovered that those DEGs were mainly involved in proliferation and angiogenesis, such as VEGF and VEGFR signaling pathway. By constructing PPI network and identifying the hub genes, we obtained three hub proteins, including PODN, ZNF521, and CFI. These three genes are closely related to the clinical prognosis of common malignant tumors, including breast cancer, ovarian cancer, lung cancer, and gastric cancer, based on KM online platform. In addition, three hub proteins have been reported less in the context of BMSCs. Moreover, the multiple roles of the three genes in common malignant neoplasms were also less studied. As illustrated in Figure 3, the upregulation of PODN, ZNF521, and CFI indicated good survival in breast cancer (Figure 3A–C) and lung cancer (Figure 3G–I), compared with the downregulation of three hub genes. Downregulation of PODN and ZNF521 indicated a good outcome in ovarian cancer (Figure 3D and E), while the upregulation of CFI indicated a good survival in ovarian cancer (Figure 3F). Besides, the downregulation of PODN and ZNF521 indicated a good outcome in gastric cancer (Figure 3J and K), while the upregulation of CFI indicated a good survival in gastric cancer (Figure 3L). As we know, malignant tumor is a group of diseases with high heterology. Moreover, a gene whether it is a tumor promoter or inhibitor depends on the microenvironment of the tumor.

However, we have noticed that the upregulation of PODN, ZNF521, and CFI indicated good survival in breast cancer (Figure 3A–C) and lung cancer (Figure 3G–I), compared with the downregulation of three hub genes. We make a bold assumption: the three hub genes may be used in the clinic for breast and lung cancers.

PODN was a seed gene in the first key modules. PODN, which is an important component of the extracellular matrix protein, belongs to the small leucine-rich repeat protein family.29 It was reported that PODN may bind type 1 collagen, thereby reducing the cell growth and migration in kidney. It is unfortunate that no literature has reported the role of PODN in tumors. ZNF521 was a seed gene in the second modules. ZNF521 is a zinc finger protein that consists almost entirely of 30 C2H2 Kruppel-like zinc fingers. Moreover, ZNF521 has been characterized as a potent inhibitor of EBF1 and is emerging as a potentially relevant contributor to the development of B-cell leukemias.31 ZNF521 contributes to clonogenic growth, migration, and tumorigenicity in medulloblastoma cells through the recruitment of the NuRD complex.30 Furthermore, ZNF521 sustains the differentiation block in myeloid-lymphoid leukemia-rearranged acute myeloid leukemia. CFI was a seed gene in the fourth modules. CFI is a serum protease that inhibits all complement pathways. The common variant rs10033900 near the CFI gene is associated with age-related macular degeneration risk in Han Chinese population. Based on the gene expression profile, we found that upregulated CFI indicates good prognosis, while Okroj et al suggested that high expression of CFI is associated with poor prognosis and recurrence in breast cancer.35
In our study, we systematically characterized DEGs from the exosomes derived from BMSCs and identified certain hub proteins and their PPI networks. Taken together, our findings suggested that three genes may be potential prognostic markers in common malignant tumors. However, the study results were obtained based on bioinformatics, which only previously considered datasets. However, further studies are required with large samples and for experimental verification.

**Figure 2** Through PPI network construction and hub module analysis identifies four central modules of key proteins.

**Notes:** (A) Protein–protein interaction network of the selected DEGs. (B–E) The four key gene modules calculated by MCODE software from the protein–protein interaction network. The red circle stands for the hub genes/proteins.

**Abbreviations:** DEGs, differentially expressed genes; MCODE, Molecular Complex Detection.
Figure 3 The three hub of the survival analysis for cancer datasets by Kaplan–Meier.

Notes: (A–C) Hub genes/proteins (PODN, ZNF521, CFI) of the survival analysis for breast cancer dataset by Kaplan–Meier. (D–F) Hub genes/proteins (PODN, ZNF521, CFI) of the survival analysis for ovarian cancer dataset by Kaplan–Meier. (G–I) Hub genes/proteins (PODN, ZNF521, CFI) of the survival analysis for lung cancer dataset by Kaplan–Meier. (J–L) Hub genes/proteins (PODN, ZNF521, CFI) of the survival analysis for gastric cancer dataset by Kaplan–Meier.

Abbreviation: HR, hazard ratio.
Acknowledgment
The authors would like to thank Hai-Ying Ma for his useful suggestions during this work.

Author contributions
All authors contributed to data analysis, drafting and revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Disclosure
The authors report no conflicts of interest in this work.

References